04/03/17

Mathématiques Statistique :

Stat descriptive : Données collecté → représentation graph
(pop taille reduite)
exemple : hystogramme
à l'aide de - moyenne
- ecart type etc ...

Stat mathématique : Extrapolation avec les résultats obtenus dans la population reduite des informations de la population totale.

Ex : Predire pourcentages de votes :
population : la France
p le % de vote pour le candidat C.
soit N la taille de la population
et n la taille de l'echantillon
→ etude statistique descriptive sur l'echantillon

$$x_i = \begin{cases} 1 & \text{si vote} \\ 0 & \text{sinon} \end{cases}$$

$$\boxed{\text{moyenne} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$ (moyenne observée)

proportion observée = moyenne.

Definition : Moyenne (cas general)

$$\boxed{\bar{x} = \sum \frac{n_i x_i}{n}}$$ avec $\sum n_i = n$ et n_i effectif associé a la valeur x_i

Variance observé : $\boxed{s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$
$VAR(X) = \sigma^2$

ecart type $\boxed{\sigma = \sqrt{s^2} = \sqrt{\sum^{n} (\ldots - \bar{x})^2}}$ avec $\sum n_i = n$

Théorème d'Hygens :

$$\boxed{\begin{aligned} s^2 &= \frac{\sum (x_i)^2}{n} - (\bar{x})^2 \\ s^2 &= \frac{\sum n_i x_i^2}{n} - (\bar{x})^2 \end{aligned}}$$

Ex :

notes x_i	12	15	8	18	9
effectif n_i	3	4	6	3	4
x_i^2	144	225	64	324	81

$$\bar{x} = \frac{3 \times 12 + 15 \times 4 + 8 \times 6 + 18 \times 3 + 9 \times 4}{20}$$

$$s^2 = \frac{3 \times 144 + 4 \times 225 + 6 \times 64 + 3 \times 324 + 4 \times 81}{20}$$

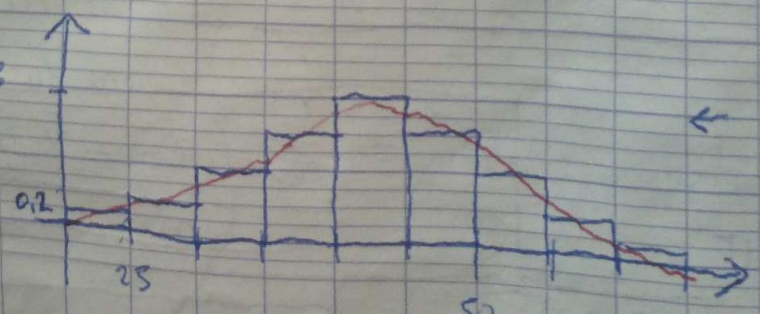
et ensuite $s^2 = \frac{\sum n_i x_i^2}{n} - (\bar{x})^2$

Coefficient de symetrie :

On appelle classe les intervalles $[x_i, x_{i+1}]$

On appelle frequences les valeurs associées $\boxed{f_i = \frac{n_i}{n}}$

Ex :



← On veut s'assurer que c'est une distribution gaussienne

On calcule le coefficient de symetrie (skwiness)

$$\cancel{\gamma_1 = \frac{\mu_3}{\sigma^2}}$$

et

$$\mu_3 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{n}$$

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

Coefficient d'applatissement

$$\gamma_2 = \frac{\mu_4}{\sigma^4}$$

$$\mu_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$

Théorème :

Si $\gamma_1 \simeq 0$ et $\gamma_2 \simeq 3$

↳ loi gaussien moyene m et ecartype σ

1er quartil valeur correspond à $\frac{1}{4}$ de la population

x_i	8	9	12	15	18
n_i	3	6	5	4	2
cumulé	3	9	14	18	20

1er Quartil $Q_1 = 8$

médiane $Q_2 = 9$

3em Quartil $Q_3 = 14$

Regression linéaire : Correlation entre 2 serie stat

Exemple : le lien entre surface appartement en fct du prix.

On trace d'abord un graphe représentant la série avec nuage de point :

Avec x_i et y_i on trouve une fonction polynomiale de degré n tel que :

$$f(x_i) = y_i \qquad \forall i \in \{1 \to n\}$$

avec $f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$

↳ impossible

<u>Méthode des moindres carrés</u> :

On cherche $\hat{y} = b_0 + b_1 X$

tel que $S = \sum_{i=1}^{n} (b_0 + b_1 X_i - y_i)^2$ soit le plus petit possible.

<u>Rappel</u> :

Soit $f : \mathbb{R}^2 \to \mathbb{R}$

$(x, y) \longmapsto f(x, y)$

On dit que (x_0, y_0) est extremum ssi :

$$\boxed{\frac{\partial f}{\partial x}(x_0, y_0) = 0 \qquad \frac{\partial f}{\partial y}(x_0, y_0) = 0}$$

Ex : $f(x, y) = 2x^2 + 3xy$

$\frac{\partial f}{\partial x} = 4x + 3y \qquad \frac{\partial f}{\partial y} = 3x \quad \Rightarrow \begin{cases} 4x + 3y = 0 \\ 3x = 0 \end{cases}$

$$\boxed{n\, S_{xy} + n\,\bar{x}\,\bar{y} = \sum x_i y_i}$$

On reprend les calculs dans ②

$$n\bar{x}\bar{y} - nb(\bar{x})^2 + b\left(n\, S_x^2 + n(\bar{x})^2\right) - \left(n S_{xy} + n\bar{x}\bar{y}\right) = 0$$

$$nb\, S_x^2 - n\, S_{xy} = 0$$

$$\boxed{b = \frac{S_{xy}}{S_x^2}} \qquad \boxed{a = \bar{y} - \frac{S_{xy}}{S_x^2}\,\bar{x}}$$

Exercice 1

1/ déjà traitée (cours)

2/ quelques notations

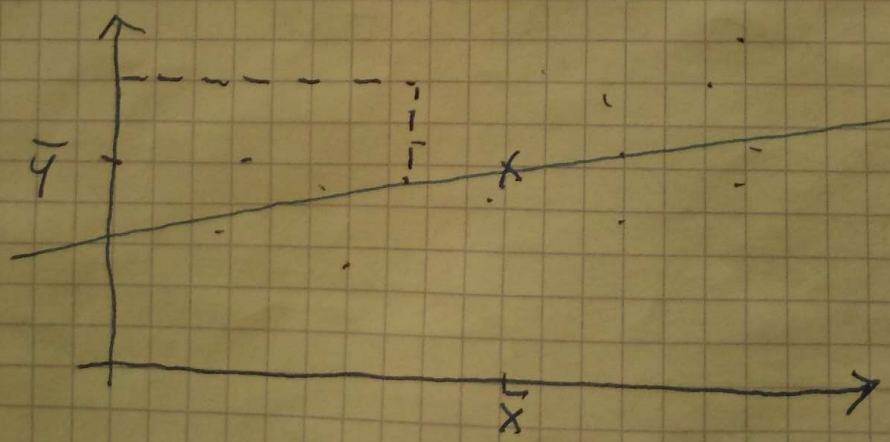
covariance

$$S_{x,y}^2 = S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} \boxed{= \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\bar{y}}$$

$$S_x^2 = S_{x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x})^2$$

$$S_y^2 = S_{y^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} = \sum_{i=1}^{n} y_i^2 - \bar{y}^2$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$



somme des ecarts carrés minimum

$$S = \sum_{i=1}^{n}(y_i - a x_i - b)^2$$

a et b sont obtenu lorsque S est minimum $\frac{\partial S}{\partial a} = 0$ $\frac{\partial S}{\partial b}$

b_0 et b_1 doit réaliser le minimum de

$$S(a;b) = \sum_{i=1}^{n} (a + bx_1 - y)$$

$$\frac{\partial S(a,b)}{\partial a} = 2\sum_{i=1}^{n} (a + bx_i - y_i)$$

$$\frac{\partial S(a,b)}{\partial b} = 2\sum_{i=1}^{n} x_i(a + bx_i - y_i)$$

$$\left.\begin{array}{l} \dfrac{\partial S(a,b)}{\partial a} = 0 \\ \\ \dfrac{\partial S(a,b)}{\partial b} = 0 \end{array}\right. \Rightarrow \begin{cases} \sum_{i=1}^{n} a + bx_i - \sum_{i=1}^{n} y = 0 \quad (1) \\ \sum_{i=1}^{n} x_i + b\sum x_i^2 - \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} xy = 0 \quad (2) \end{cases}$$

$$(1) \Rightarrow \sum_{i=1}^{n} a + b\overbrace{\sum_{i=1}^{n} x_i}^{n\bar{x}} - \overbrace{\sum_{i=1}^{n} y_i}^{n\bar{y}} = 0$$

$$na + bn\bar{x} - n\bar{y} = 0$$

$$n(a + b\bar{x} - \bar{y}) = 0$$

$$\bar{y} = a + b\bar{x} \quad \text{donc} \quad a = \bar{y} - b\bar{x}$$

(2) on a :

$$(\bar{y} - b\bar{x})n\bar{x} + b\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i = 0$$

On note

$$S_{x^2} = \frac{\sum x_i^2}{n} - (\bar{x})^2$$

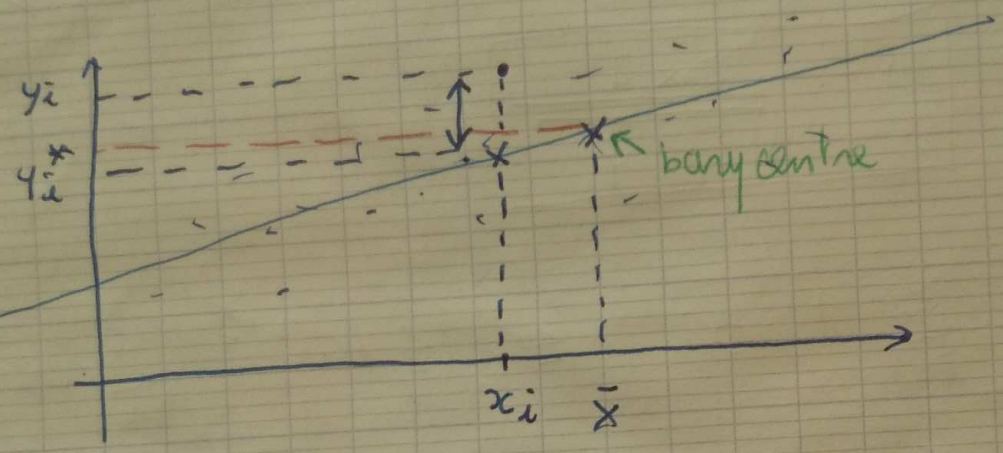
Donc :

$$\boxed{n S_{x^2} + n(\bar{x})^2 = \sum_{i=1}^{n} x_i^2}$$

⇒ la covariance : $S_{xy} = \dfrac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\bar{y}$

/ On note $y_i^* = \dfrac{\sum_{i=1}^{n} (y_i - y_i^*)^2}{n}$

Montrer que : $S^2_{x,y} = S_y^2 (1 - r^2)$



$$\boxed{S^2_{yx} = \frac{\sum_{i=1}^{n} (y_i - y_i^*)^2}{n}}$$

$$\boxed{S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}}$$

$$\boxed{S^2_{yx} = S_y^2 (1 - r^2)}$$

$(x_i, y_i^*) \in$ droite de regression
(x_i, y_i) un point du nuage de point
$(\bar{x}, \bar{y})$ le barycentre du nuage.

Preuve : $n S^2_{xy} = \sum_{i=1}^{n} (y_i - y_i^*)^2$

$$= \sum_{i=1}^{n} y_i^2 - 2 y_i y_i^* + y_i^{*2}$$

$$= \sum_{i=1}^{n} y_i^2 - 2 \sum_{i=1}^{n} y_i y_i^* + \sum_{i=1}^{n} y_i^{*2}$$

$$S_y^2 = \frac{\sum y_i^2}{n} - (\bar{y})^2 \Rightarrow \sum_{i=1}^{n} y_i^2 = n S_y^2 + (\bar{y})^2$$

$$\sum_{i=1}^{n} y_i y_i^* = \sum_{i=1}^{n} y_i (a x_i + b)$$

$$= a \sum_{i=1}^{n} x_i y_i + b \sum_{i=1}^{n} y_i$$

$$S_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\bar{y}$$

$$\sum_{i=1}^{n} x_i y_i = n S_{xy} + \bar{x}\bar{y}$$

$$\sum_{i=1}^{n} y_i^{*2} = \sum_{i=1}^{n} (a x_i + b)^2$$

$$= \sum_{i=1}^{n} (a^2 x_i^2 + 2ab x_i + b^2)$$

$$= a^2 \sum_{i=1}^{n} x_i^2 + 2ab \sum_{i=1}^{n} x_i + b^2 \sum_{i=1}^{n} 1$$

$$= a^2 (n S_x^2 + n(\bar{x})^2) + 2abn\bar{x} + nb^2$$

$$n S_{yx}^2 = n S_y^2 + n(\bar{y})^2 - 2[an S_{xy} + n\bar{x}\bar{y} + bn\bar{y}]$$
$$+ a^2[n S_x^2 + n(\bar{x})^2] + 2ab n\bar{x} + nb^2$$

Or $a = \frac{S_{xy}}{S_x^2}$ et $b = \bar{y} - a\bar{x}$

$$n S_{xy}^2 = n S_y^2 + n(\bar{y})^2 - 2\left[n \frac{S_{xy}^2}{S_x^2} + n\bar{x}\bar{y}\right.$$
$$\left. + n\left(\bar{y} - \frac{S_{xy}}{S_x^2}\right)\bar{y}\right] + \frac{S_{xy}^2}{S_x^2} \times [n S_x^2 + n(\bar{x})^2]$$
$$+ 2n \frac{S_{xy}}{S_x^2}\left(\bar{y} - \frac{S_{xy}}{S_x^2}\right)\bar{x} + n\left(\bar{y} - \frac{S_{xy}}{S_x^2}\right)^2$$

En utilisant le fait que :

$$\boxed{r^2 = \frac{S_{xy}^2}{S_x^2 S_y^2}}$$

on obtient :

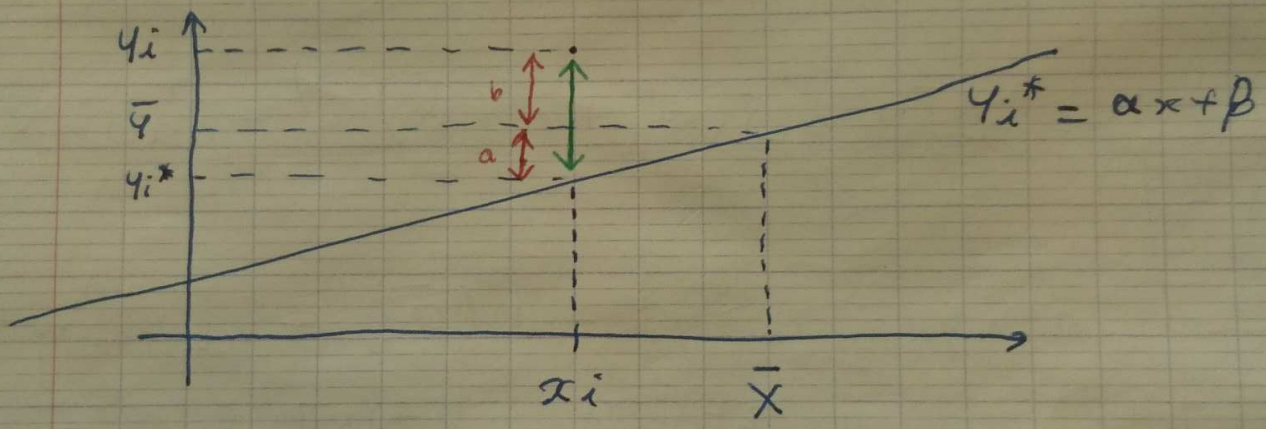
$$\boxed{\quad}$$

Le coefficient de détermination linéaire :

Il mesure la qualité de l'ajustement par le modèle de la régression linéaire :

$$\boxed{R^2 = \frac{\sum_{i=1}^{n} (y_i^* - \bar{y})^2}{\sum (y_i - \bar{y})^2}}$$

Propriété : $\boxed{0 \leq R^2 \leq 1}$



$$R^2 = \frac{\sum a_i^2}{\sum b_i^2}$$

$$S_{yx}^2 = \frac{\sum (y_i - y_i^*)^2}{n} = \frac{\sum [(y - \bar{y}) + (\bar{y} - y_i^*)]^2}{n}$$

$$= \frac{\sum (y_i - \bar{y})^2 - \sum (y_i^* - \bar{y})^2}{n}$$

$$= S_y^2 - \frac{\sum (y_i^* - \bar{y})^2}{n}$$

r^2 = indépendance $X \Leftrightarrow Y$

R^2 = qualité de l'ajustement par la régression linéaire

d'autre part : $R^2 = \dfrac{\frac{1}{n} \sum (y_i^* - \bar{y})^2}{S_y^2}$

donc : $\dfrac{\sum (y_i^* - \bar{y})}{n} = S_y^2 R^2$

$$S_{yx}^2 = S_y^2 - S_y^2 R^2 = S_y^2 (1 - R^2)$$

et on a montré que : $S_{yx}^2 = S_y^2 (1 - r^2)$

$$\boxed{r^2 = R^2}$$

18/03/17

Preuve :

$$\sum (y_i - \bar{y}) + \sum (\bar{y} - y_i^*)^2 + 2 \sum (y_i - \bar{y})(\bar{y} - y_i^*)$$

$$= -2 \sum (\bar{y} - y_i^*)^2$$

Il suffit de montrer :

$$\sum (y_i - \bar{y})(\bar{y} - y_i^*) = - \sum (\bar{y} - y_i^*)^2$$

$$\sum \left[\left((y_i - y_i^*) + (y_i^* - \bar{y}) \right) \times (\bar{y} - y_i^*) \right]$$

$$= \sum (y_i - y_i^*)(\bar{y} - y_i^*) + \sum (y_i^* - \bar{y})(\bar{y} - y_i^*) - \sum (y_i^* - \bar{y})^2$$

Ex 3 :

$$y = n_0 e^{-ct}$$

à $t = 1$ $\quad n_1 = n_0 e^{-c}$

par approximation

$$\frac{n_1}{n_0} = e^{-c} \qquad \ln\left(\frac{n_1}{n_0}\right) = -c \qquad \boxed{c = \ln\left(\frac{n_0}{n_1}\right)}$$

2) On cherche n_0 et c tel que $y = n_0 e^{-ct}$

$$\ln(y) = \underbrace{\ln(n_0)}_{a} - \underbrace{ct}_{b}$$

$$z = a + bt$$

t	0	1	3	4	5	6
y	200	130	48	30	18	11
$\ln y$	5,3	4,9	3,9	3,4	2,9	2,39

$$b = \frac{3,4 - 2,9}{4 - 5} = -0,5$$

or $\ln(n_0) = a = 200$

$b = -c = 0,5$

Statistique inférentielle : Estimation.

$P(X_1 = 0 \cap X_2 = 1) = P(X_1 = 0) \cdot P(X_2 = 1) = p(1-p)$
$= f(p)$

Maximum $\Rightarrow f'(p) = 1 - 2p = 0$

$\Rightarrow p = 1/2$

Comment appréhender statistiquement un problème ?

- On a n observations (données) $x_1 \dots x_n$ d'un phenomène ou experience.
- On decide de concidérer les données comme realisation de n variables aléatoires independant et identiquement distribuées selon la même loi de parametre θ.

Il reste donc à estimer la vrai valeur du parametre inconnu θ ou a tester une hypothese pour cela, on doit definir :

- Qu'est ce qu'une estimation ?
- Comment le construir ? avec quelle methode ?
- Comment evaluer sa qualité ?

Premier Cas de Bernouilli :

On a $x_1 \dots x_n \in \{0 ; 1\}$
On suppose $X_1 \dots X_n$ des variable independante identiquement distribuées. $\boxed{\text{v. a. i. i}}$

Selon la loi de bernouilli de parametre p inconnu.

On detinit la fonction de vraissamblance.

$f(p, x_1, x_2, \dots x_n) = P(X_1 = x_1, X_2 = x_2 \dots$

(intersection) = SUM $\cap$

Les X_i $(i = 1, \dots n)$ étant indépendants :

$$f(p, x_1 \dots x_n) = P\left(\bigcap_{i=1}^{n} (X_i = x_i)\right)$$

= SUM $\times$ (produit)

$$= \prod_{i=1}^{n} P(X_i = x_i)$$

or $P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}$

On conclut que $f(p, x_1 \dots x_n) = \prod_{i=1}^{n} p^{x_i} (1-x)^{1-x_i}$

Chercher le maximum de $f(p, x_1, \dots x_n)$ revi à chercher le maximum de $\log(f(p, x_1 \dots x_n))$

f admet un maximum en x_0 :

$\Rightarrow f'(x_0) = 0$ et $f''(x_0) \leq 0$

$(\log(f(x)))' = f'/f \Rightarrow (\log(f(x)))'' = \dfrac{f'' \times f - f}{f^2}$

$= f''/f - f$

$(\log f(x))''(x_0) \leq 0 \Leftrightarrow f''(x_0) \leq 0$

Donc $\log f(p, x_1 \dots x_p) = \log \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$

$$= \sum_{i=1}^{n} \log(p^{x_i}(1-p)^{1-x}$$

$$= \sum_{i=1}^{n} \log p^{x_i} + \log(1-p)^{1-x}$$

$$= \sum_{i=1}^{n} x_i \log p + (1-x_i) \log (1-p)$$

$$= \sum_{i=1}^{n} x_i \log p + \sum_{i=1}^{n} (1-x_i) \log (1-p)$$

$$= \log p \underbrace{\sum_{i=1}^{n} x_i}_{= \bar{x} n} + \log (1-p) \underbrace{\sum_{i=1}^{n} (1-x_i)}_{n - n\bar{x}}$$

$$\sum_{i=1}^{n} 1 - \sum_{i=1}^{n} x_i$$

$$= n\bar{x} \log p + n(1-\bar{x}) \log (1-p)$$

$$\frac{\partial \log f(p, x_i \dots x_n)}{\partial p} = \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p}$$

On résoud l'equation.

$$\frac{\partial}{\partial p} \log f(p, x_1 \dots x_n) = 0$$

$$\frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p} = 0$$

$$\frac{n\bar{x}(1-p) - n(1-\bar{x})p}{p(1-p)} = 0$$

$$\Rightarrow n\bar{x}(1-p) - n(1-\bar{x})p = 0$$

$$n\bar{x} - n\bar{x}p - np - n\bar{x}p = 0$$

$$n\bar{x} - np = 0 \Rightarrow \boxed{p = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

Definition : Soit $X_1 \dots X_n$ une suite de v.a.i selon la m̂ loi de parametre θ
On appelle estimation de θ la fonction $h : \mathbb{R}^n \to \mathbb{R}$ tel que $h(x_1, \dots x_n) = \theta$

Exemple : dans le cas de la loi de bernouilli :
la fonction h est donné par $h(X_1 \dots X_n) = \frac{\sum X_1}{n}$
et donc une estimation de p est la variable aléatoire : $\bar{X} = h(x_1 \dots x_n) = \frac{\sum x_1}{n}$

Definition : On dit qu'une estimation est sans biais lorsque $E(h(X_1 \dots X_n)) = \theta$

Exemple : Pour bernouilli
$E[\bar{X}] = E\left[\frac{\sum X_i}{n}\right] = \frac{1}{n} \sum E[X_i] = \frac{1}{n} \sum p$
donc $\bar{X}$ est sans biais $\to \frac{np}{n} = p$

Definition : On dit qu'un estimateur est convergent ssi $\lim_{n \to \infty} V(h(x_1, \dots x_n)) = 0$

Exemple : $V(\bar{X}) = V\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i)$
or $V(X_i) = p(1-p) \quad \forall i \in \{1 \dots n\}$
donc $V(\bar{X}) = \frac{1}{n^2} \sum p(1-p) = \frac{1}{n^2} np(1-p)$

$= p(1-p) \times 1/n \xrightarrow[n \to +\infty]{} 0$

donc $\lim_{n \to \infty} V(\bar{X}) = 0$

Demarche à suivre : 1 On considère la V.A. X_i

$f(\theta, x_1 \dots x_n) = \prod f_{x_i}$

où f_{x_i} est la densité de probabilité associée à X_i

③ On evalue la fonction $\log f(\theta, x_1 \dots x_n)$
$= \sum \log f_{x_i}(p, x_1 \dots x_n)$

et on calcule $\frac{\partial}{\partial \theta} \log f(\theta x_1 \dots x_n)$

④ On résoud l'equation $\frac{\partial}{\partial \theta} \log f(\theta, x_1 \dots x_n) = 0$

le qui nous donne une estimation de θ

⑤ On verifie que cet estimateur est sans biais

⑥ On verifie que cet estimateur est convergent.

Loi de poisson $P(\lambda)$

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k}$$

le probleme est de trouver un estimateur de λ
Soient $x_1 \dots x_n$ les valeurs observées qui sont la realisation des VA ii $X_1, \dots X_n$

$$f(\lambda, x_1 \dots x_n) = \prod P(X_i = x_i) = \prod \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$\log f(\lambda, x_1 \dots x_n) = \log \prod \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$= \sum \log \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = \sum \left(\log \frac{\lambda^{x_i}}{} + \log e^{-\lambda} \right)$$

$i=1$

$$\frac{\partial}{\partial\lambda} \log f(\lambda, x_1 \dots x_n) = \frac{\sum x_i}{\lambda} - n$$

$$= \frac{n\bar{X}}{\lambda} - n$$

$$\frac{n\bar{X}}{\lambda} - n = 0 \iff \frac{\bar{X}}{\lambda} - 1 = 0$$

$$\iff \lambda = \bar{X} = \frac{\sum x_i}{n}$$

d'où un estimateur de λ est donné par $\bar{X} = \frac{\sum X_i}{n}$

$$E(\bar{X}) = E\left(\frac{\sum x_i}{n}\right) = \frac{n\lambda}{n} = \lambda$$

car $E(X) = \lambda$

Donc $\bar{X}$ est un estimateur sans biais de λ

$V(\bar{X}) = V\left(\frac{\sum x_i}{n}\right) = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$ d'où $\lim_{n \to +\infty} V(\bar{X}) = 0$

donc l'estimateur est convergeant.

Loi exponentielle

$$f_X(t) = \lambda e^{-\lambda t}$$

$$f(X, x_1 \dots x_n) = \prod f_{x_i}(x_i) = \prod \lambda e^{-\lambda x_i}$$

$$= \lambda^n e^{-\lambda \sum x_i}$$

$$= \lambda^n e^{-n\lambda\bar{X}}$$

$$\log f(\lambda, x_1 \dots x_n) = \log \lambda^n e^{-n\lambda\bar{X}}$$

$$= n \log \lambda - n\lambda\bar{X}$$

$$\frac{\partial}{\partial\lambda} f(\lambda, x_1 \dots x_n) = \frac{n}{\lambda} - n\bar{X}$$

$\Rightarrow \frac{1}{\lambda} = \bar{X} \quad \Rightarrow \quad \lambda = \frac{1}{\bar{X}}$

donc une estimation de λ est donné par

$$\frac{1}{\bar{X}} = \frac{n}{\sum X_i}$$

$$E(\bar{X}) = E\left[\frac{\sum X_i}{n}\right] = \frac{n}{n} E[X] = E[X] = \frac{1}{\lambda}$$

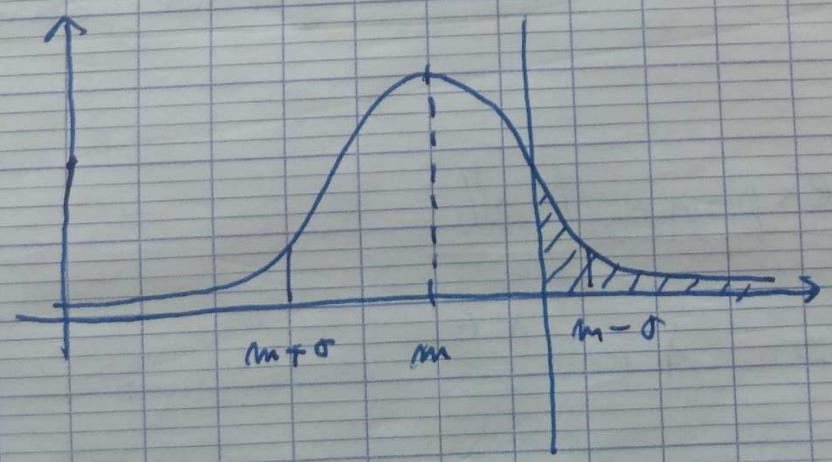
il s'agit bien d'un estimateur sans biais

de meme $V(\bar{X}) = \frac{V(X)}{n} = \frac{1}{\lambda^2 n}$

Loi Normal $N(m, \sigma^2)$

1) Si $X \sim N(m, \sigma^2)$ alors la densité de probabilité est donnée par $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$

$P(X \in]m-\sigma, m+\sigma[) = 0{,}95$:



2) $f : \mathbb{R}^2 \mapsto \mathbb{R}$

$(x, y) \mapsto f(x, y)$

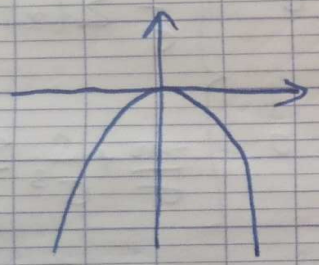
alors f admet en (x_0, y_0) un extremum ssi

$$\partial f(x_0, y_0) = \partial f(x_0, y_0) = 0$$

De plus f admet en (x_0, y_0) un maximum ssi

$$0 \geqslant \det(\text{Hess } f(x_0, y_0)) = \begin{vmatrix} \frac{\partial^2 f}{\partial x^2}(x_0, y_0) & \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) \\ \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) & \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \end{vmatrix}$$

Savoir si il y a un extremum :



$f(x) = -x^2$

$f'(x) = -2x$ $\qquad f'(x) = 0 \Rightarrow x = 0$

$f''(x) = -2$ $\qquad f''(0) = -2 \leq 0$

Soient $x_1 \dots x_n$ les resultats des données observées

Soient $X_1 \dots X_n$ une suite de vaii selon la loi normal de param m et σ

La fonction du maximum de vraissemblance est donnée par :

$$f(m, \sigma, x_1, \dots x_n) = P(X_1 = x_1 \dots X_n = x_n)$$
$$= \prod_{i=1}^{n} f_{X_i}(x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i - m)^2}{2\sigma^2}} = \frac{1}{2\pi^{n/2}\sigma^n} e^{-\sum_{i=1}^{n} \frac{(x_i - m)^2}{2\sigma^2}}$$

$$\log f(m, \sigma, x_1, \dots x_n) = \log\left[\frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\sum \frac{(x_i - m)^2}{2\sigma^2}}\right]$$

$$= \log\left[\frac{1}{(2\pi)^{n/2}\sigma^n}\right] + -\frac{1}{2\sigma^2}\sum (x_i - m)^2$$

$$= -n \log(2\pi) - n \log\sigma - \frac{1}{\ldots}\sum (x_i - m)^2$$

On calcule :

$$\frac{\partial}{\partial m}\left[\log f(m, \sigma, x_1 \dots x_n)\right] = \frac{1}{\sigma^2}\sum(x_i - m)$$

$$\frac{\partial}{\partial \sigma}\left[\log f(m, \sigma, x_1 \dots x_n)\right] = \frac{-m}{\sigma} + \frac{1}{\sigma^3}\sum(x_i - m)^2$$

On resoud :

$$\begin{cases} \frac{\partial}{\partial m}\log\left[f(m, \sigma, x_1 \dots x_n)\right] = 0 \\ \frac{\partial}{\partial \sigma}\log\left[f(m, \sigma, x_1 \dots x_n)\right] = 0 \end{cases} \quad \begin{cases} \frac{1}{\sigma}\sum(x_i - m) = 0 \\ \frac{-m}{\sigma} + \frac{1}{\sigma^3}\sum(x_i - m)^2 = 0 \end{cases}$$

$$\begin{cases} (1)\ \sum x_i - \sum m = 0 \\ (2)\ \frac{1}{\sigma}\left[-n + \frac{1}{\sigma^2}\sum(x_i - m)^2\right. = 0 \end{cases}$$

$$(1) \Rightarrow n\bar{X} - nm = 0 \Rightarrow \boxed{m = \bar{X}}$$

$$\boxed{\Rightarrow \sigma^2 = \frac{\sum(x_i - m)^2}{n}}$$

Conclusion :

* Un estimateur de m est donné par $\bar{X} = \frac{\sum X_i}{n}$
* Un estimateur de σ^2 est donné par $S^2 = \frac{\sum(X_i - \bar{X})^2}{n}$

Question : Estimateur sans biais ?

$$E[\bar{X}] = E\left[\frac{\sum X_i}{n}\right] = \frac{1}{n}\sum \underbrace{E[X_i]}_{m}$$

$$= \frac{nm}{n} = m$$

Donc $\bar{X}$ est un estimateur sans biais de m

Ce n'est pas le cas de S^2

$$E[S^2] = \frac{n-1-\sigma^2}{n}$$

Pour prendre cet estimateur sans biais on le remplace par :

$$\tilde{S}^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

$$E[\tilde{S}^2] = \sigma^2$$

Pourquoi $\bar{x}$ et s^2 realisent le maximum pour $\log[f(m, \sigma, x_1 \dots x_n)]$

$$\frac{\partial}{\partial m} \log f(m, \sigma, x_1 \dots x_n) = \frac{1}{\sigma^2} \sum (x_i - m)$$

$$\frac{\partial^2}{\partial m^2} \log f(m, \sigma, x_1 \dots x_n) = \frac{-1}{\sigma^2} \sum 1 = \frac{-n}{\sigma^2}$$

$$\frac{\partial^2}{\partial \sigma \partial m} \log f(m, \sigma, x_1 \dots x_n) = \frac{\partial}{\partial \sigma}\left[\frac{1}{\sigma^2} \sum (x_i - m)\right]$$

$$= -\frac{2}{\sigma^3} \sum (x_i - m) = 0$$

Conclusion :

$$\frac{\partial}{\partial \sigma} \log f(m, \sigma, x_1 \dots x_n) = \frac{-n}{\sigma} + \frac{1}{\sigma^3} \sum (x_i - m)^2$$

$$\frac{\partial^2}{\partial \sigma^2} \log f(m, \sigma, x_1 \dots x_n) = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum (x_i - m)^2$$

$$= \frac{n}{\sigma^2} - \frac{3}{\sigma^2} \frac{n}{\sigma^2} \frac{\sum (x_i - m)^2}{n}$$

$$= \frac{n}{\sigma^2} - \frac{3n}{\sigma^2} = \frac{-2n}{\sigma^2}$$

$$\mathrm{Hess}(\log f(m, \sigma, x_1 \dots x_n)) = \begin{pmatrix} \frac{-n}{\sigma^2} & 0 \\ 0 & \frac{-2n}{\sigma^2} \end{pmatrix}$$

et $\det(\mathrm{Hess}(\log f(m, \sigma, x_1 \dots x_n))) = \frac{2n^2}{\sigma^4}$

Interval de Confiance

1) Motivation de l'etude :

"Etes vous favorable à une constitution européenne ?"
lors d'une enquête 171 / 300 disent "oui".

On considère $x_1 \dots x_n$ les resultats de l'enquête

$$x_1 \begin{cases} 1 & \text{si oui} \\ 0 & \text{si non} \end{cases}$$

On considère $X_1 \dots X_n$ une suite de VAii de loi de Bernouilli de paramètre p.

Le problème revient à chercher un intervalle I tel que

$P(p \in I) = 1 - \alpha$

$1 - \alpha$ niveau de confiance

α risque

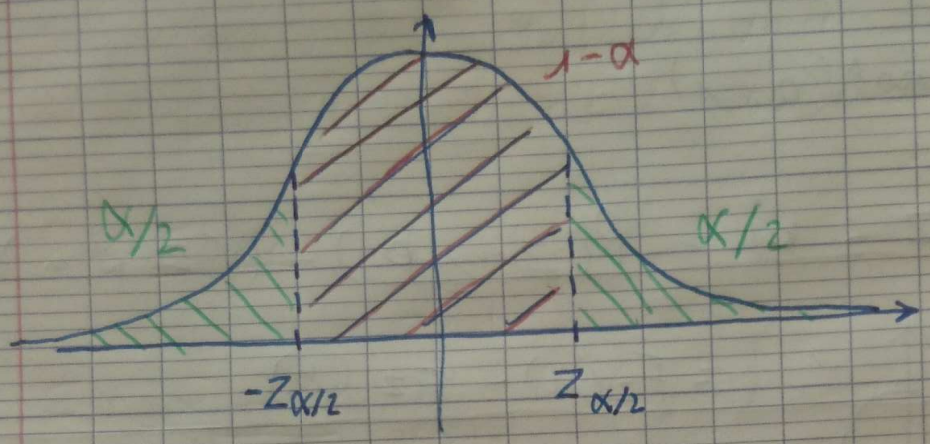
La théorie d'estimation permet de donner une estimation de p puis de $\bar{X} = \frac{\sum X_i}{n}$

$E[\bar{X}] = p$

$V(\bar{X}) = \frac{p(1-p)}{n}$

Le théorème centrale limite permet d'affirmer

$$Z = \frac{\bar{X} - E[\bar{X}]}{\sqrt{V(\bar{X})}} \sim N(0,1)$$



$$P(|Z| < Z_{\alpha/2}) = 1-\alpha$$

$$|Z| \leq Z_{\alpha/2} \Rightarrow -Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$$

$$\Rightarrow -Z_{\alpha/2} \leq \frac{\bar{X} - E[\bar{X}]}{\sqrt{V(\bar{X})}} \leq Z_{\alpha/2}$$

$$\Rightarrow -Z_{\alpha/2} \leq \frac{\bar{X} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \leq Z_{\alpha/2}$$

$$\Rightarrow -Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leq \bar{X} - p \leq Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Shulski: $Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \sim N(0,1)$ avec $\bar{p} = \frac{\sum x_i}{n}$

$$\Rightarrow -Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leq p - \bar{X} \leq Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

donc $\bar{X} - Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \leq p \leq \bar{X} + Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

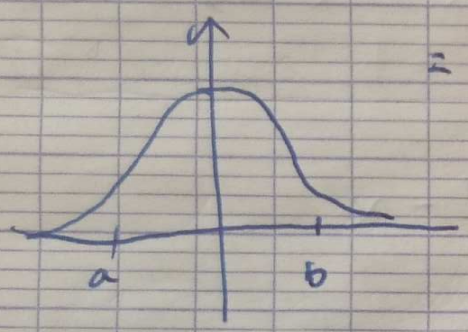
estimation
observée

Rappel

loi $Z \sim N(0;1)$

$$P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$$

$$= F_Z(a) - F_Z(b)$$



$$F_Z(-a) = 1 - F_Z(a)$$

On choisit $\alpha = 0{,}05$ $\quad \alpha/2 = 0{,}025$

$1 - \alpha/2 = 0{,}975$

$Z_{\alpha/2}$? $\quad P(|Z| < Z_{\alpha/2}) = 1 - \alpha$

$$\Rightarrow P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$$

$$\Rightarrow F_Z(Z_{\alpha/2}) - F(-Z_{\alpha/2}) = 1 - \alpha$$

$$\Rightarrow 2F(Z_{\alpha/2}) - 1 = 1 - \alpha$$

$$\Rightarrow 2F(Z_{\alpha/2}) = 2 - \alpha$$

$$\Rightarrow 2F(Z_{\alpha/2}) = 1 - \alpha/2$$

Par lecture de table de loi normal

$Z_{\alpha/2} = 1{,}96$

$$I = \left[\frac{171}{300} - 1{,}96 \times \sqrt{\frac{\frac{171}{300} \cdot \frac{129}{300}}{300}} \; ; \; \frac{171}{300} + 1{,}96 \sqrt{\frac{\frac{171}{300} \cdot \frac{129}{300}}{}} \right]$$

Rappel

$P(p \in I_c) = 1-\alpha$ $\quad$ α = risque

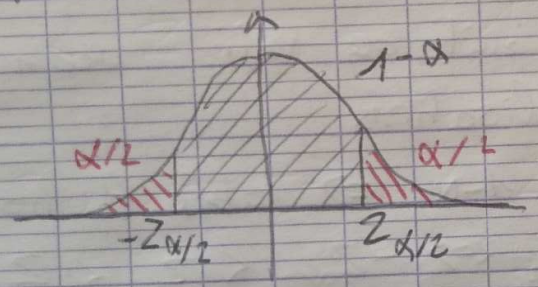
$1-\alpha$ = confiance

I_c intervalle confiance

Soient $X_1 \dots X_n$ suite de v.a.i suivant la loi de bernouilli et $x_1 \dots x_n$ les realisation

$$\bar{x} = \frac{\sum X_i}{n}$$

$$Z = \frac{|\bar{x} - p|}{\sqrt{V(\bar{x})}} \underset{n \to +\infty}{\overset{T.C.L}{\sim}} N(0,1)$$



$$P(|Z| < Z_{\alpha/2}) = 1-\alpha$$

à partir d'un risque α qu'on se donne, on calcule $1-\alpha$ et par lecture de la table statistique de la loi normal on obtient $Z_{\alpha/2}$ donc avec un niveau de confiance $= 1-\alpha$

On a :

$$|Z| < Z_{\alpha/2} \iff -Z_{\alpha/2} < Z < Z_{\alpha/2}$$

$$-Z_{\alpha/2} < \frac{\bar{x} - p}{\sqrt{V(\bar{x})}} < Z_{\alpha/2}$$

$$-Z_{\alpha/2} < \frac{p - \bar{x}}{\sqrt{V(\bar{x})}} < Z_{\alpha/2}$$

$$\bar{x} - Z_{\alpha/2}\sqrt{V(\bar{x})} < p < \bar{x} + Z_{\alpha/2}\sqrt{V(\bar{x})}$$

$$V(\bar{x}) = V\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} \sum V(X_i)$$

$$= \frac{1}{n^2} \sum p(1-p)$$

$$= \frac{p(1-p)}{n}$$

2. Cas de la loi Normal

Intervalle de confiance lorsque σ est connu

Soient $X_1 \dots X_n$ suite de vaii selon la loi $N(m, \sigma)$
et $x_1 \dots x_n$ réalisations
Avec α on calcule $1-\alpha$ => table loi normal => $Z_{\alpha/2}$
On a $|Z| < Z_{\alpha/2} \Leftrightarrow -Z_{\alpha/2} < Z < Z_{\alpha/2}$

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < m < \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

et donc $I_c = \left[\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$

Exercice 1

a) σ et m connue

$$I_c = \left[\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

Il s'agit d'une situation ou la loi étudier est la loi normal de parametre μ et d'écart type $\sigma = 0{,}03$
On est donc dans le cas de recherche d'intervalle de confiance de la moyenne avec ecart type connu de la loi normal
L'intervalle de confiance est

$$I_c = \left[\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

avec $\bar{x} = 89{,}2$ $n = 16$ $\sigma = 0{,}3$
il reste donc à determiner $Z_{\alpha/2}$
or $P(|Z|$

e 7 et 8

Un contrôle regression linéaire. samedi le 13/05

Intervalle de confiance
et
Test d'hypothèses

Exercice 1 :

Considérons une population admettant la distribution $N(\mu,\sigma^2)$. On y prélève un échantillon aléatoire simple d'effectif 16 qui donne une moyenne $\overline{x}=83,2$. Déterminer un intervalle de confiance pour μ au niveau de confiance égale à 0,95 dans les cas suivants :

a) σ^2 est connue, avec $\sigma^2=0,09$.

b) σ^2 est inconnue, et on note un écart type observé de $s=0.3$.

Exercice 2 :

On pèse 10 palettes de briques de la même fabrication ; et on a obtenu les résultats suivants (en kilogrammes) :

759 751 755 756 761 766 770 752 760 767

1) a- Calculer la moyenne m (moyenne empirique) de cet échantillon à 10^{-1}.

b- Rappeler l'expression de l'estimateur non biaisé de la variance;

En déduire la valeur de la variance σ^2 à 10^{-1}.

On suppose que la distribution des masses de palettes de briques suit la loi normale $N(m,\sigma)$.

2) Déterminer un intervalle de confiance au risque 1% pour la moyenne.

3) Déterminer un intervalle de confiance au risque 2% pour l'écart type.

4) On suppose que l'écart type σ est *parfaitement connu* : $\sigma^2=42$.

Déterminer le nombre n de palettes de briques, pour obtenir un intervalle de confiance de longueur 0,22 de la moyenne au risque 6%.

Exercice 3 :

Supposons que les notes (sur un total de 200) obtenues lors d'un examen de statistiques soient distribuées suivant une loi normale. On examine un échantillon de 26 notes qui fournit une moyenne de 125 et un écart type de 10.

a) Déterminer un intervalle de confiance pour μ un intervalle de confiance pour σ^2 au niveau de confiance égale à 0,95.

b) A quel niveau de probabilité α peut-on dire que la moyenne des notes se situe entre 120 et 130 avec une probabilité égale à $1-\alpha$.

déterminer interval de confiance de la variance

Exercice 4 :

On désire tester l'hypothèse selon laquelle le poids moyen des individus d'une population vaut 65 Kg. Sachant que l'écart type est de 4 Kg, et qu'on prélève un échantillon de d'effectif 36, déterminer avec un risque $\alpha=0,05$:

a) Les valeurs de la moyenne observée qui $\overline{x}$, qui conduisent au rejet de cette hypothèse.

b) La probabilité β de commettre l'erreur de seconde espèce, sil la population a réellement la moyenne de 66 Kg.

Exercice 5 :

22/04/2017

$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = F_Z(Z_{\alpha/2}) - F(-Z_{\alpha/2})$

$= F_Z(Z_{\alpha/2}) - (1 - F_Z(Z_{\alpha/2}))$

$= 2F_Z(Z_{\alpha/2}) - 1$

Or $1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05$

$\alpha/2 = 0{,}025$

$1-\alpha/2 = 0{,}975$

On cherche donc $Z_{\alpha/2}$ tel que

$P(Z < Z_{\alpha/2}) = 0{,}975$

En utilisant la table statistique on obtient

$Z_{\alpha/2} = 1{,}96$

$IC = \left[83{,}2 - \frac{1{,}96 \times 0{,}3}{4} ; 83 + \frac{1{,}96 \times 0{,}3}{4}\right]$

$= [83{,}05 ; 83{,}35]$

(b) σ est inconnu

$X_1 \dots X_n$ vaii et $x_1 \dots x_n$ realisations $N(m, \sigma)$

$\frac{\bar{X} - m}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - m}{\sigma/\sqrt{n}}$ lorsque σ est inconnu on ne peut plus utiliser le théorème centrale limite $\boxed{TCL}$.

On donne un estimateur sans biais de σ^2

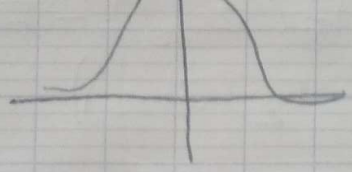
$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$

$T = \frac{\bar{X} - m}{\frac{S}{\sqrt{n}}} \sim T_{\alpha/2, n-1}$ En poursuivant la m̂ démarche on obtient :

$IC = \left[\bar{X} - t_{\alpha/2, n-1}\frac{S}{\sqrt{n}} ; \bar{X} + t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}\right]$

$\bar{x} = 83{,}2$ (moy observé)

$S = 0{,}3$ (ecart type observé) $P(T < t_{\frac{\alpha}{2}; n-1}) = 1 - \frac{\alpha}{2}$



		0,975
1		⋮
⋮		⋮
15	. . . .	2,1315

Ici ecart type σ est inconnu

determiner intervalle confiance moyenne.

donc $IC = [\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} ; \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}]$

Or pour $n-1 = 15$ et $\alpha = 0,05$

$t_{\alpha/2, n-1} = 2,131$

donc $IC = [83,2 - 2,131 \times \frac{0,3}{4} ; 83,2 + 2,131 \times \frac{0,3}{4}]$

$= [83,04 ; 83,36]$

~~Ex~~

Ex3

une loi N avec equart type connu

intervalle de confiance donné par

$IC = [\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$

$= [125 - 1,96 \times \frac{10}{\sqrt{25}} ; 125 + 1,96 \times \frac{10}{\sqrt{25}}]$

$= [121,16 ; 128,84]$

b) On cherche α tel que

$P(120 < \bar{X} < 130) = 1 - \alpha$

$P\left(\frac{120 - 125}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X} - 125}{\frac{\sigma}{\sqrt{n}}} < \frac{130 - 125}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$

$P(-5 < \ldots < 5) = \ldots$

protected

$$2 F_Z\left(\frac{S_{10}}{\sqrt{26}}\right) - 1 = 1 - \alpha$$

$$1 - \frac{\alpha}{2} = F_Z\left(\frac{\sqrt{26}}{2}\right)$$

$$1 - \frac{\alpha}{2} \simeq F_Z(2,5)$$

$$1 - \frac{\alpha}{2} = 0,9938 \Rightarrow \frac{\alpha}{2} = 0,0062$$

$$\alpha = 0,0124$$

```
public String
```

```
pain Choc (bad crouskillat)
this. crouskillat = crouskillat;
toString() return "est" + roi;
```

```
Veinoisaure
String toString()
return "Veinoise = " + nom
```

super.toString()

Rappel :

1 - Cas d'une proportion

$$\boxed{I_c = \left[\bar{p} - Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{\sqrt{n}}} \; ; \; \bar{p} + Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{\sqrt{n}}}\right]}$$

avec $F_Z(Z_{\alpha/2}) = 1 - \alpha/2$

2 - Cas de la moyenne d'une loi normale

- σ est connu :

$$\boxed{I_c = \left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \; ; \; \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

avec $\boxed{F_Z(Z_{\alpha/2}) = 1 - \alpha/2}$

- σ inconnu

$$\boxed{I_c = \left[\bar{x} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} \; ; \; \bar{x} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]}$$

3ème cas de la variance σ^2 d'une loi normale

- moyenne m connue

$$\boxed{I_c = \left[\frac{nS^2}{c_2} \; , \; \frac{nS^2}{c_1}\right]}$$

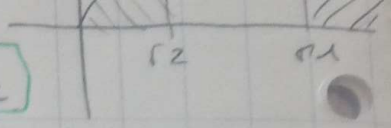
n la taille de l'échantillon

$$\boxed{S^2 = \frac{\sum (x_i - m)^2}{n}}$$

avec S^2 la variance observé dans la population observé

$$\Downarrow$$

$$\boxed{P(x < r_2) = 1 - \alpha/2}$$



- moyenne m inconnue

$$\boxed{S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}} \quad \boxed{\bar{x} = \frac{\sum x_i}{n}}$$

$$\boxed{I_c = \left[\frac{(n-1)S^2}{r^2} ; \frac{(n-1)S^2}{r_1} \right]}$$

r_1 et r_2 donnés par table de loi du χ^2 à $n-1$ degré de liberté

Exercice 3

$\sigma = 10$

a) ~~calcule moyenne observé = 125 = $\bar{x}$~~

$1-\alpha = 0,95$

Il s'agit de determiner l'intervalle de confiance de la variance σ^2 avec moyenne m inconnu.

$S^2 = 10^2 = 100 \quad n = 26$

$$I_c = \left[\frac{(n-1)S^2}{r^2} ; \frac{(n-1)S^2}{r_1} \right]$$

$P(x > r_2) = 0,025 \Rightarrow r_2 = 40,65$

$P(x < r_1) = 0,975 \Rightarrow r_1 = 13,12$

Interval de confiance de variance :

$$I_c = \left[\frac{25 \times 100}{40,65} ; \frac{25 \times 100}{13,12} \right] = [61,5 ; 190,5]$$

Interval de equart type

$I_c = [7,8 ; 13,8]$

Exercice 2

a) calcule moyenne observé $\bar{x} = 759,7$

b) $S^2 = \frac{\sum (X_i - \bar{x})^2}{n-1}$ estimateur de variance

$= \frac{\sum (x_i - \bar{x})}{9}$ $\quad S^2 = 41,34$

$S = 6,43$

Math 9 stat
29/04/17

Ex2

2) Interval de confiance de la moyenne avec $\alpha = 0{,}01$
σ étant inconnu, donc

$$Ic = \left[\bar{x} - t_{n-1,\alpha/2} \frac{S}{\sqrt{n}} ; \bar{x} + t_{n-1,\alpha/2} \frac{S}{\sqrt{n}} \right]$$

$\bar{x} = 759{,}7$ kg $\quad S = 6{,}43 \quad n = 10$
$\alpha = 0{,}01 \quad \alpha/2 = 0{,}005$

$$P(T < t_{9;0{,}005}) = 1 - \alpha/2 = 0{,}995$$

$$\Rightarrow t_{9;0{,}005} = 3{,}25$$

Intervalle de confiance

$$Ic = \left[759{,}7 - 3{,}25 \times \frac{6{,}43}{\sqrt{10}} ; 759{,}7 + 3{,}25 \times \frac{6{,}43}{\sqrt{10}} \right]$$

$$= [753{,}1 ; 766{,}3]$$

3) Même chose avec $\alpha = 0{,}02$ determiner ecart type : avec moyenne théorique inconnu.
donc l'interval de confiance est donné par :

$$Ic = \left[(n-1) \frac{S^2}{r2} ; (n-1) \frac{S^2}{r1} \right]$$

r_1 et r_2 determiner avec table χ^2 à $n-1$ degrés de libertés

ici $n = 10$ donc $n - 1 = 9$
$\alpha = 0{,}02 \Rightarrow \alpha/2 = 0{,}01$
$P(X > r_2) = 0{,}01 \quad r_2 = 21{,}67$
$P(X > r_1) = 1 - \alpha/2 = 0{,}99$
$r_1 = 2{,}08 \quad Ic = \left[\frac{9 \times 41{,}34}{21{,}67} ; \frac{9 \times 41{,}34}{2{,}08} \right]$

Ic de ecart type $\Rightarrow Ic = [4{,}1 ; 13{,}4]$

Dans ce cas l'intervalle de confiance de la moyenne est $I_c \left[\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$

la longueur de l'intervalle est :

$$2 Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Ce qui revient à chercher n tel que :

$$2 Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 0,22$$

et $P(Z < Z_{\alpha/2}) = 1 - \alpha/2$

$= 1 - \frac{0,06}{2} = 0,97$

$Z_{\alpha/2} = 1,88$

donc il faut trouver n tel que :

$$2 \times 1,88 \times \frac{\sqrt{42}}{\sqrt{n}} = 0,22$$

$$n = \frac{(2 \times 1,88)^2 \times 42}{0,22}$$

S^2

$n = 2699$

$\sigma = S$

4) Cas d'une loi quelconque

4.1) Interval de confiance de la moyenne

le TCL permet d'obtenir l'intervalle de confiance dans le cas où σ est connu

$$I_c = \left[\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right.$$

4.2) Cas où σ est inconnu $n < 30$

$$I_c = \left[\bar{x} - t_{n-1 \frac{\alpha}{2}} \frac{S}{\sqrt{n}} ; \bar{x} + t_{n-1} \frac{S}{\sqrt{n}}\right]$$

2ème cas $n \geq 30$

$I_c = [\bar{x}$

avec $Z_{\alpha/2}$ donné par la table de la loi normal

<u>Definition</u> : Précision d'une estimation
On appelle precision d'une estimation la valeur correspondant à la moitié de l'intervalle de confiance

<u>Exemple</u> : avec le cas de la moyenne avec σ connu la quantité $U = \frac{\sigma}{\sqrt{n}} \frac{Z_\alpha}{2}$ est la precision de l'estimation

* Dans le cas de la moyenne avec σ inconnu la quantité $U = \frac{S}{\sqrt{n}} t_{n-1, \alpha/2}$ est la précision de l'estimation
* Dans le cas de la proportion

$$U = Z_{\alpha/2} \frac{\sqrt{P(1-P)}}{\sqrt{n}}$$

<u>Definition</u> : taille de l'échantillon optimale la taille d'échantillon optimale est le nombre d'observ requises pour obtenir une precision U au niveau de confiance $1-\alpha$

<u>Exemple</u> : Chercher n tel que

- $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \mu$ avec $P(Z < Z_{\alpha/2}) = 1 - \alpha$
- $t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} = \mu$ avec $P(T < t_{n-1 \frac{\alpha}{2}}) = 1$
- $Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{\sqrt{n}}} = \mu$ avec $P(Z < Z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$

avec $Z_{\alpha/2}$ donné par la table de la loi normal

<u>Definition</u> : Précision d'une estimation
On appelle precision d'une estimation la valeur correspondant à la moitié de l'intervalle de confiance

<u>Exemple</u> : avec le cas de la moyenne avec σ connu la quantité $U = \frac{\sigma}{\sqrt{n}} \frac{Z_\alpha}{2}$ est la precision de l'estimation

* Dans le cas de la moyenne avec σ inconnu la quantité $U = \frac{S}{\sqrt{n}} t_{n-1, \alpha/2}$ est la précision de l'estimation
* Dans le cas de la proportion

$$U = Z_{\alpha/2} \frac{\sqrt{P(1-P)}}{\sqrt{n}}$$

<u>Definition</u> : taille de l'échantillon optimale la taille d'échantillon optimale est le nombre d'observations requises pour obtenir une precision U au niveau de confiance $1-\alpha$

<u>Exemple</u> : Chercher n tel que

- $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \mu$ avec $P(Z < Z_{\alpha/2}) = 1 -$
- $t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} = \mu$ avec $P(T < t_{n-1 \frac{\alpha}{2}}) =$
- $Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{\sqrt{n}}} = \mu$ avec $P(Z < Z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$

avec $Z_{\alpha/2}$ donné par la table de la loi normal

<u>Definition</u> : Précision d'une estimation
On appelle precision d'une estimation la valeur correspondant à la moitié de l'intervalle de confiance

<u>Exemple</u> : avec le cas de la moyenne avec σ connu la quantité $U = \frac{\sigma}{\sqrt{n}} \frac{Z_\alpha}{2}$ est la precision de l'estimation

* Dans le cas de la moyenne avec σ inconnu la quantité $U = \frac{S}{\sqrt{n}} t_{n-1, \alpha/2}$ est la précision de l'estimation
* Dans le cas de la proportion

$$U = Z_{\alpha/2} \frac{\sqrt{P(1-P)}}{\sqrt{n}}$$

<u>Definition</u> : taille de l'échantillon optimale la taille d'échantillon optimale est le nombre d'observations requises pour obtenir une precision U au niveau de confiance $1-\alpha$

<u>Exemple</u> : Chercher n tel que

- $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \mu$ avec $P(Z < Z_{\alpha/2}) = 1 -$
- $t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} = \mu$ avec $P(T < t_{n-1} \frac{\alpha}{2}) =$
- $Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{\sqrt{n}}} = \mu$ avec $P(Z < Z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$

Exercice d'application :

Compagnie veut estimer proportion passager classe affaire

n 13 passagers dont 20 en classe affaire

a) Determiner IC de proportion en classe A avec $\alpha = 0,$

b) quelle taille ~~pour~~ echantillon n pour avoir precision 0,05?